Person reidentification Part2: Hem Bahadur Gurung and Sachin Awal

Three types of loss

- Idnetity loss
  It treats the training process of person Re-ID as an image classification problem, i.e., each identity is a distinct class.
- Verification Loss.
  It optimizes the pairwise relationship, either with a contrastive loss or binary verification loss. Binary verification discriminates the positive and negative of an input image pair.
- Triplet loss. It treats the Re-ID model training process as a retrieval ranking problem. The basic idea is that the distance between the positive pair should be smaller than the negative pair by a pre-defined margin.

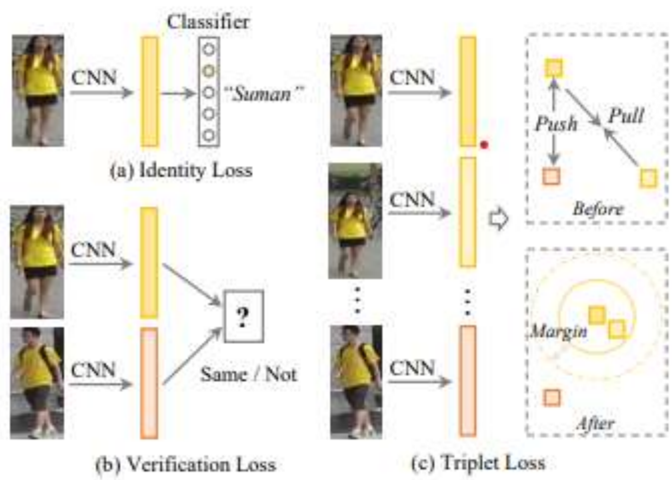


Figure: Types of loss

Process involve in person reidentification

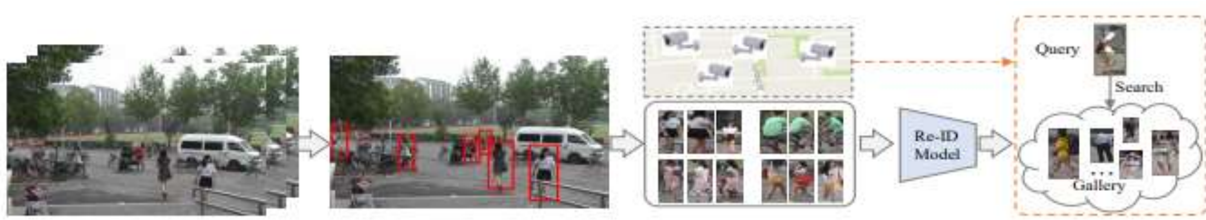


Fig: The flow of designing a practical person Re-ID system, including five main steps:

- **Step 1. Raw Data Collection from images and videos:** Access raw video data from surveillance cameras

- **Step 2. Bounding Box Generation**: Extracting the bounding boxes which contain the person images from the raw video data by the person detection or tracking algorithms.
- **Step 3. Training Data Annotation**: Annotating the cross-camera labels. Training data annotation is usually indispensable for discriminative Re-ID model learning due to the large cross-camera variations.
- **Step 4: Model Training**: Training a discriminative and robust Re-ID model with the previous annotated person. images/videos.
- **Step 5: Pedestrian Retrieval:** The testing phase conducts the pedestrian retrieval. Given a person-of-interest (query) and a gallery set, we extract the feature representations using the Re-ID model learned in previous stage.

Table Closed-world vs. Open-world Person Re-ID.

Closed-world	**Open-world**
Single-modality Data	Heterogeneous Data
Bounding Boxes Generation	Raw Images/Videos
Sufficient Annotated Data	Unavailable/Limited Labels
Correct Annotation	Noisy Annotation
Query Exists in Gallery	Open-set

Recent Research paper

Learning Person Re-Identification Models From Videos With Weak Supervision(2021)

- Supervised person Re-identification techniques, suffer from the burden of massive annotation requirement. While Unsupervised methods overcome this need for labeled data, but perform poorly compared to the supervised alternatives.
- Introduce the problem of learning person re-identification models from videos with weak supervision.
- The weak nature of the supervision arises from the requirement of video-level labels, i.e. person identities who appear in the video, in contrast to the more precise frame-level annotations.
- Propose a multiple instance attention learning framework for person re-identification using such video-level labels.
- Specifically, we first cast the video person re-identification task into a multiple instance learning setting, in which person images in a video are collected into a bag. The relations between videos with similar labels can be utilized to identify persons, on top of that, we introduce a co-person attention mechanism which mines the similarity correlations between videos with person identities in common.

- The attention weights are obtained based on all person images instead of person tracklets in a video, making our learned model less affected by noisy annotations.

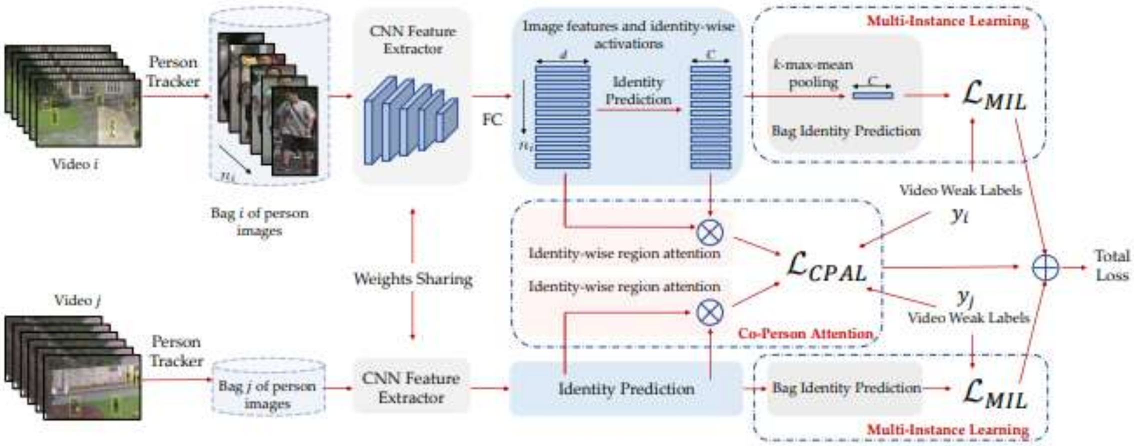


Fig: A brief illustration of our proposed multiple instance attention learning framework for video person re-id with weak supervision.

1. For each video, we group all person images obtained by pedestrian detection and tracking algorithms in a bag and use it as the inputs of our framework.
2. The bags are passed through a backbone CNN to extract features for each person image.
3. Furthermore, a fully connected (FC) layer and an identity projection layer are used to obtain identity-wise activations.
4. On top of that, the MIL loss based on k-max-mean-pooling strategy is calculated for each video.
5. For a pair of videos (i, j) with common person identities, we compute the CPAL loss by using high and low attention region for the common identity.
6. Finally, the model is optimized by jointly minimizing the two loss functions.

Gait recognition for person re-identification (2020)

I. Gait recognition, which is the recognition based on the walking style, is mostly used for this purpose due to that human gait has unique characteristics that allow recognizing a person from a distance.
II. This paper proposes a gait recognition approach for person re-identification starts with estimating the angle of the gait first, and this is then followed with the recognition process, which is performed using convolutional neural networks(CNN).
III. Herein, **multitask convolutional neural network models** and extracted **gait energy images** (GEIs) are used to estimate the angle and recognize the gait.

IV. GEIs are extracted by first detecting the moving objects, using background subtraction techniques. Training and testing phases are applied to the following three recognized datasets: CASIA-(B), OU-ISIR, and OU-MVLP.

V. The proposed method is evaluated for background modeling using the Scene Background Modeling and Initialization (SBI) dataset.

VI. The proposed gait recognition method showed an accuracy of more than 98% for almost all datasets.

VII. Results of the proposed approach showed higher accuracy compared to obtained results of other methods result for CASIA-(B) and OU-MVLP and form the best results for the OU-ISIR dataset.

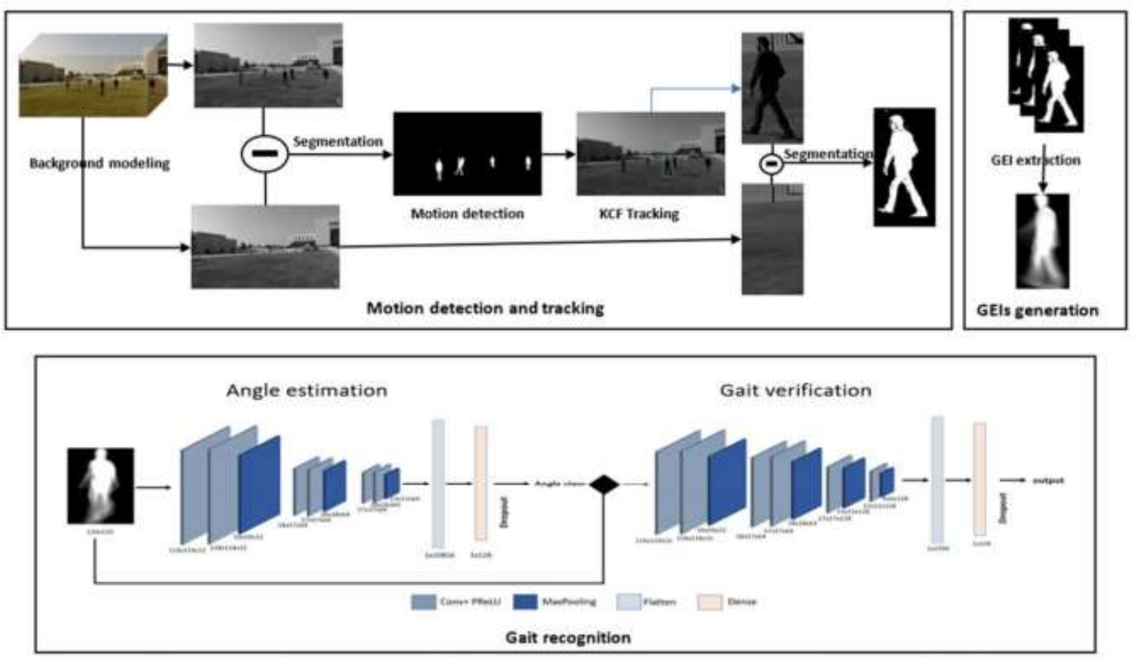


Fig: Gait recognition for person re-identification

Fine-Grained Shape-Appearance Mutual Learning for Cloth-Changing Person Re-Identification(2021)

I. Current person re-identification (Re-ID) methods largely depend on color appearance, which is not reliable when a person changes the clothes.

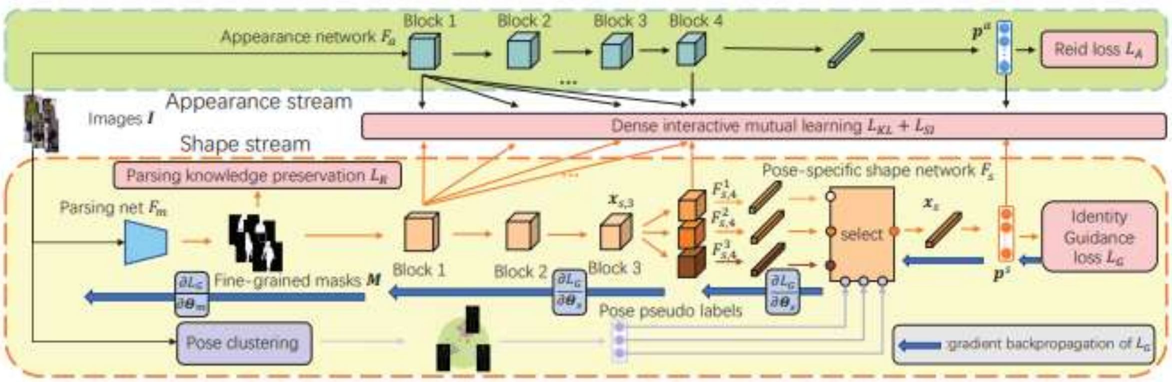
II. Cloth-changing Re-ID is challenging since pedestrian images with clothes change exhibit largen intra-class variation and small inter-class variation. Some significant features for identification are embedded in unobvious body shape differences across pedestrians.

III. To explore such body shape cues for cloth-changing Re-ID,
Propose a Fine-grained Shape-Appearance Mutual learning framework (FSAM), a two-stream framework that learns fine-grained discriminative body shape knowledge in a shape stream and transfers it to an appearance stream to complement the cloth-unrelated knowledge in the appearance features.

IV. Specifically, in the shape stream, FSAM learns fine-grained discriminative mask with the guidance of identities and extracts fine-grained body shape features by a pose-specific multi-branch network.

V. To complement cloth unrelated shape knowledge in the appearance stream, dense interactive mutual learning is performed across low-level and high-level features to transfer knowledge from shape stream to appearance stream, which enables the appearance stream to be deployed independently without extra computation for mask estimation.

VI. We evaluated our method on benchmark cloth-changing Re-ID datasets and achieved the start-of-the-art performance.



Person Re-identification using Heterogeneous Local Graph Attention Networks(2021)

I. Recently, some methods have focused on learning local relation among parts of pedestrian images for person reidentification (Re-ID), as it offers powerful representation capabilities only provide the intra-local relation among parts within single pedestrian image and ignore the inter-local relation among parts from different images, which results in incomplete local relation information.

II. Propose a novel deep graph model named Heterogeneous Local Graph Attention Networks (HLGAT) to model the inter-local relation and the intra-local relation in the completed local graph, simultaneously.

III. Specifically, First construct the completed local graph using local features, and we resort to the attention mechanism to aggregate the local features in the learning process of inter-local relation and intra-local relation so as to emphasize the importance of different local features.

IV. As for the inter-local relation, propose the attention regularization loss to constrain the attention weights based on the identities of local features in order to describe the inter-local relation accurately.

V. As for the intra-local relation, propose to inject the contextual information into the attention weights to consider structure information.

VI. Extensive experiments on Market-1501, CUHK03, DukeMTMC-reID and MSMT17 demonstrate that the proposed HLGAT outperforms the state-of-the-art methods.

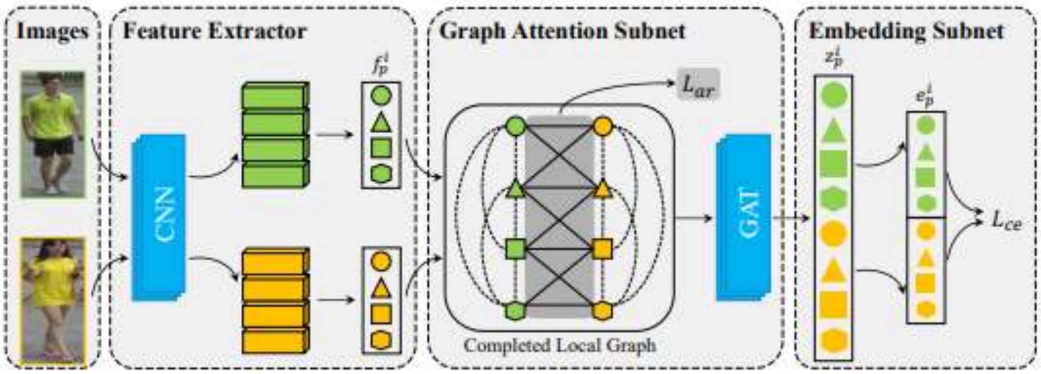


Figure: Pipeline of the proposed HLGAT. As for the completed local graph, the solid lines indicate the inter-local edges, and the dotted lines represent the intra-local edges

Overview

Feature Extractor.

- Feed pedestrian images into CNN to obtain feature maps, and then we adopt uniform partition strategy to split these feature maps into several horizontal grids.
- Finally, we extract local features using the global max pooling operation on these grids.

Graph Attention Subnet.

- Regard the local features as the nodes to construct the completed local graph to learn the inter-local relation and the intra-local relation, simultaneously.
- These nodes are linked by the inter-local edges and the intra-local edges.
- For each node in the graph, weight it's all neighbor nodes using the attention weights, and then aggregate them to obtain the two kinds of relations.
- Meanwhile, constrain the attention weights of inter-local edges using the attention regularization loss to describe the interlocal relation accurately, and we inject the contextual information into the attention weights of intra-local edges to consider structure information.

• **Embedding Subnet**.

- In this subnet, we apply independent fully connected (FC) layers to reduce the dimension of the features extracted from Graph Attention Subnet.
- We utilize these dimension-reduced features as the final features, and make identity prediction on them.

Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification (2021)

i. This paper considers the problem of unsupervised person re-identification (re-ID), which aims to learn discriminative models with unlabeled data.
ii. One popular method is to obtain pseudo-label by clustering and use them to optimize the model.
iii. Although this kind of approach has shown promising accuracy, it is hampered by 1) noisy labels produced by clustering and 2) feature variations caused by camera shift. The former will lead to incorrect optimization and thus hinders the model accuracy.
iv. The latter will result in assigning the intra-class samples of different cameras to different pseudo-label, making the model sensitive to camera variations.
v. Concretely, propose a Dynamic and Symmetric Cross-Entropy loss (DSCE) to deal with noisy samples and a camera-aware meta-learning algorithm (MetaCam) to adapt camera shift.
vi. DSCE can alleviate the negative effects of noisy samples and accommodate the change of clusters after each clustering step.
vii. MetaCam simulates cross-camera constraint by splitting the training data into meta-train and meta-test based on camera IDs.
viii. With the interacted gradient from meta-train and meta-test, the model is enforced to learn camera-invariant features.
ix. Extensive experiments on three re-ID benchmarks show the effectiveness and the complementary of the proposed DSCE and Metacam. Our method outperforms the state-of-the-art methods on both fully unsupervised re-ID and unsupervised domain adaptive re-ID.

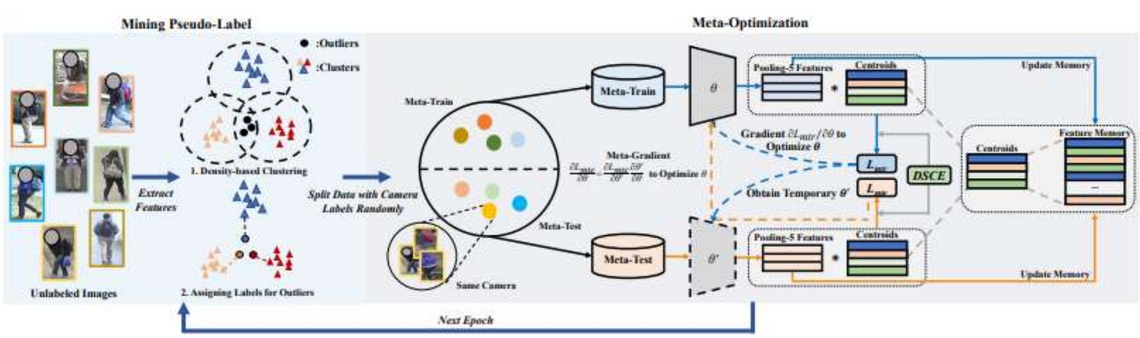


Fig: The framework for unsupervised re-ID, includes two training stages, i.e., "Mining Pseudo-Label" stage and "Met Optimization".

- The first stage assigns samples with pseudo-label based on DBSCAN.
- The second stage splits the training data into meta-train and meta-test sets based on camera labels and optimizes the model with the proposed meta-learning strategy.
- This camera-aware meta-learning (Metacam) encourages the model to learn camera-invariant features.
- To reduce the negative impact of noisy labels, also propose a dynamic and symmetric cross-entropy loss (DSCE) that is used for both meta-train and meta-test data.

- The memory module saves the features of all samples, which enables us to dynamically build class centroids and thus to be adaptable to the change of clusters.

Discover Cross-Modality Nuances for Visible-Infrared Person Re-Identification(2021)

I. Visible-infrared person re-identification (Re-ID) aims to match the pedestrian images of the same identity from different modalities.

II. Existing works mainly focus on alleviating the modality discrepancy by aligning the distributions of features from different modalities.

III. However, nuanced but discriminative information, such as glasses, shoes, and the length of clothes, has not been fully explored, especially in the infrared modality.

IV. Without discovering nuances, it is challenging to match pedestrians across modalities using modality alignment solely, which inevitably reduces feature distinctiveness.

V. Propose a joint Modality and Pattern Alignment Network (MPANet) to discover cross-modality nuances in different patterns for visible infrared person Re-ID, which introduces a modality alleviation module(MAM) and a pattern alignment module(PAM) to jointly extract discriminative features.

VI. Specifically, first **propose a modality alleviation module (MAM)** to dislodge the modality information from the extracted feature maps.

VII. Then, devise a **pattern alignment module (PAM)**, which generates multiple pattern maps for the diverse patterns of a person, to discover nuances.

VIII. Finally, introduce a mutual mean learning fashion to alleviate the modality discrepancy and propose a center cluster loss to guide both identity learning and nuances discovering.

IX. Extensive experiments on the public SYSU-MM01 and RegDB datasets demonstrate the superiority of MPANet over state-of-the-arts.

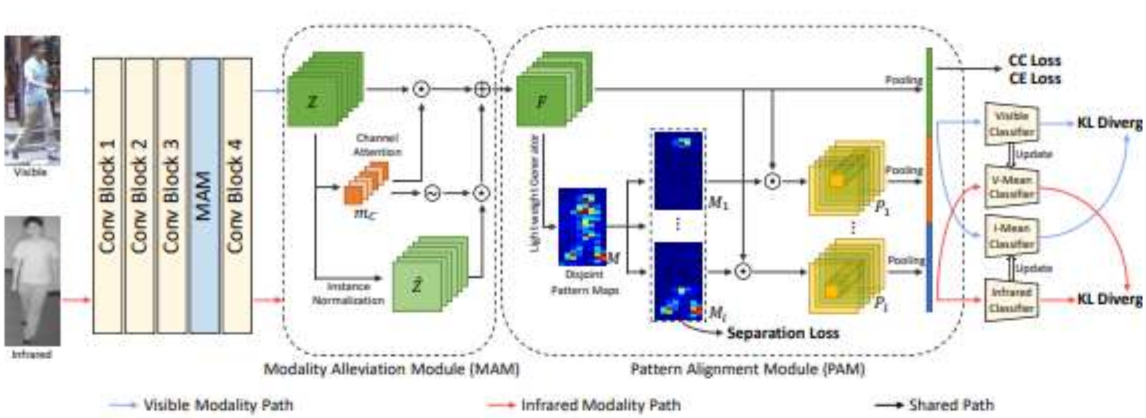


Fig: Framework of the proposed joint Modality and Pattern Alignment Network (MPANet).